DOCUMENT RESUME

ED 059 545	08	EA 004 083

AUTHOR	Wasik, John L.
TITLE	A Review and Critical Analysis of Mathematical Models Used for Estimating Enrollments in Educational Systems.
INSTITUTION	North Carolina State Univ., Raleigh. Center for Occupational Education.
SPONS AGENCY	National Center for Educational Research and Development (DHEW/OE), Washington, D.C.
BUREAU NO	BR-7-0348
PUB DATE	71
GRANT	OEG-2-7-070348-2698
NOTE	28p.; Center Technical Paper No. 5
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Community Colleges; *Enrollment Projections; Enrollment Rate; Enrollment Trends; *Mathematical Models; Multiple Regression Analysis; Planning; *Statistical Analysis

ABSTRACT

This report examines existing mathematical methods that have been utilized to develop educational enrollment estimates and analyzes the applicability of these methods in the community college setting. Models surveyed are based on (1) extrapolation--by either survival cohort or cohort-regression, (2) structural flow, and (3) Markov-type procedures. The paper then presents some of the theoretical and practical problems associated with the formulation and implementation of these models. The final chapter presents a rationale for developing enrollment estimates by institution or curriculum in community colleges. (Author)

A REVIEW AND CRITICAL ANALYSIS OF MATHEMATICAL MODELS USED FOR ESTIMATING ENROLLMENTS IN EDUCATIONAL SYSTEMS

John L. Wasik
Departments of Statistics and Psychology
North Carolina State University at Raleigh

****

The research reported herein was performed pursuant to a contract with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

****

Center Technical Paper No. 5

CENTER FOR OCCUPATIONAL EDUCATION
North Carolina State University at Raleigh
Raleigh, North Carolina

1971

Project No. BR 7-0348

Grant No. OEG-2-7-070348-2698

PREFACE

As enrollments climb in other areas of postsecondary education, community colleges also face rising numbers of students. It seems reasonable to assume that these enrollments will continue to increase, and educational institutions must plan ahead for facilities and personnel to serve them. Planning ahead requires some accurate estimate of the expected enrollment increase, and several models for this estimation have been developed.

This report examines closely the existing models and analyzes their applicability in the community college setting.

The Center wishes to thank Dr. Wasik for his work in preparing this enlightening report; Mrs. Sue King for editing the manuscript; and the entire Center staff for their efforts toward the publication of this technical paper.

John K. Coster
Director

SUMMARY

The utilization of mathematical models for educational enrollment projection has a brief but rapidly developing history. This paper describes mathematical methods that have been utilized to develop educational enrollment estimates. Models surveyed are based upon (1) extrapolation--by either survival cohort or cohort-regression, (2) structural flow, and (3) Markov-type procedures. The paper then presents some of the theoretical and practical problems associated with the formulation and implementation of these models. The final chapter presents a rationale for developing enrollment estimates by institution or curriculum in community colleges.

TABLE OF CONTENTS

INTRODUCTION

The use of planning procedures in education has a rather short history. The year 1959 has been generally accepted as the beginning of attempts to apply systems analysis and model building to educational phenomena. This beginning was marked by the publication by the Rand Corporation of the monograph Systems Analysis in Education by Joseph Kershaw and Roland McKean. Since that start there has been an increasing use of mathematical modeling and simulation procedures to study educational problems. The wide attention accorded the conferences sponsored by the Organization for Economic Cooperation and Development (OECD) on Mathematical Models and Educational Planning in 1966 and the 1967 Symposium on Operations Analysis in Education sponsored by the United States Office of Education indicates that a substantial amount of interest exists among educators for using mathematical models in solving educational problems.

One area in which the outputs of modeling procedures are particularly applicable is the development of future demands for educational services. The usefulness of these mathematical modeling procedures has been recognized and commented upon; however, there has been little in the way of reports of the use of such procedures to capture such demand factors as student flow through an educational system.

One of the educational specialty areas in which mathematical modeling would seem to have a great deal to offer is the administrative sphere. Present interest in the use of a Programming Planning and Budgeting (PPB) system for the effective allocation of available resources has been evident at both the national and local levels. Real data or accurate estimates of such program elements as course demand or institutional development are required for efficient use of PPB models. Since most of these types of program elements will be related to the size of institutional enrollment, an important factor in developing a PPB system is an accurate estimate of enrollment.

In many instances, institutional development requires that decisions be made years before the expected enrollment increases occur. This is especially true in planning for the development of curricula which require extensive and often costly laboratory layouts. Actions must be initiated at the earliest possible time to ensure that new facilities and additional instructional personnel will be available when the enrollment demands them. The commitment of the state-wide educational systems to utilizing educational planning procedures increases the need for methods to accurately predict future enrollment trends.

SCHEMES FOR CLASSIFYING EDUCATIONAL PLANNING MODELS

Two different classification schemes have been reported for the purpose of categorizing models developed specifically for educational planning. The scheme developed by Koulourianos (1967) considers two broad classifications of educational planning models: paediometric and economic. Within this scheme, models concerned with the dynamics of education occurring without any influence from the economy as a whole are referred to as paediometric, while economic models are specifically concerned with the dynamic interrelationships of education and the economy. Superimposed upon and orthogonal to this two-level classification scheme is a third way of categorizing models according to whether the model assumptions are based upon rate of return, manpower planning, or social demand. This classification of social models is based upon the premise that education is more or less exogenous to the economic system and the determination of the demand for educated workers. Thus, there is no requirement that the educational system indicate how these demands are to be met.

Another classification procedure for classifying educational models was presented in a paper developed for the Educational Policy Research and Support Center of the System Development Corporation (Wurtele, 1967). This scheme used three categories to classify educational planning models. Wurtele classified models according to whether they represented (1) the education system or some of its components, (2) education as one of the components in the economy, or (3) the technology of the educational process of learning.

A comparison of the two classificatory schemes indicates that Koulourianos and Wurtele were categorizing models according to subject rather than to the structure of the model itself. As will be suggested later in this paper, model structure, which is concerned with different approaches to modeling the phenomenon of interest, can also be considered as providing a scheme under which the specific modeling attempts in education can be classified.

It is also obvious that the two first classifications of the two categorization procedures refer to the same types of models. However, as noted, Wurtele considers a third kind of model which is concerned with identified psychological learning processes, while Koulourianos, an economist, utilizes a cross-classification scheme of interest to economists.

This paper will be concerned with models of the first type. There are two reasons for focusing on the educational system alone. First, the paper is being written from the viewpoint of the professional educator who is interested in providing administrators with effective tools for planning. Second, it has not been demonstrated whether growth in the education sector precedes or lags behind the economic sector; therefore, until organized attempts are made to interrelate

recruitment into various curricula as a function of demand, there appears to be no real reason for including economic elements in the development of educational planning procedures. The models discussed in this paper can also be referred to as demographic models since they are specially designed for use in the projection of student enrollment in educational systems. The following sections will review, in turn, methods for obtaining enrollment estimates by models utilizing extrapolation, structural, or Markov-type processes. In addition, the strengths and weaknesses of the use of the three types of models for developing enrollment projections will be discussed. A strategy will also be suggested for developing enrollment estimates in diverse educational states, such as curricula in a community college.

Extrapolation Procedures

Survival-Ratio Method

The procedure that is most widely used to project student enrollment is referred to as the survival-ratio method. Variants of this procedure have been described by Adair (1969) and Cross and Sederberg (1968) for use in the development of enrollment projections. Both of these methods use cohorts to develop student survival-ratios. The cohort method can be described as an accounting of the whereabouts of similar groups of individuals throughout their educational careers. Since the individual is exposed to the educational process for an extended time period, it is obvious that the cohort method is particularly appealing to those wishing to obtain stable estimates of student transition from one program to another.

Adair (1969) suggests calculating a survival rate for a grade i by finding the ratio of number of students in the i^{th} grade at time t as compared to the number of students in grade $i - 1$ at time $t - 1$. These survival rates are calculated for a period of years and then averaged to provide mean survival ratios. This procedure is carried out for each of the grades so that a projection for the number of students in grade i at time t is found by multiplying the number of students in grade $i - 1$ at time $t - 1$ by the average survival ratio for grade i. This procedure suffers somewhat from the limitation that all data are based upon actual enrollment data and, thus, significant birth rate or migration trends will not be picked up until one year after the deviant group has entered the school at grade 1 or, for postsecondary education, as first-year students.

Cross and Sederberg (1969) utilized a computerized model to give Minnesota school district administrators estimates of future school enrollment in their school district by grades. Their model also used cohort survival-rate information. Numbers of students entering the elementary school for the first time were estimated from census estimates of the number of children of ages 1 - 4 years residing in census tracts served by a particular school. It would be expected that this model would provide more accurate enrollment estimates than the Adair

model since it utilized information on children of pre-school age who would likely attend school at age six. Output from the computer program developed to carry out the calculations provided users with patterns of survival ratios that were observed to be operating the previous ten years and a series of enrollment projections based on the averages of annual survival rates.

Cohort-Regression Methods

Two methods were developed to predict the numbers of public high school graduates in the state of Minnesota as part of a methodological study on prediction of educational attendance conducted by the Department of Statistics of the University of Minnesota (Brown and Savage, 1960). A live-birth method was concerned with developing a simple regression of the number of graduates on the number of live births 18 years previously. A cohort method was also developed which used a regression technique to predict numbers of students in a particular grade by utilizing indices of students' tendency to pass from grade to grade as transition proportions, or numbers of live births six years earlier to estimate enrollment in the first grade.

For both models, the numbers of births observed in year t - 18 is multiplied by either the regression coefficient associated with ratio of graduates at time t to births at time t - 18 (Model I) or by the product of the transition rates of from one grade to another and from the twelfth grade to graduation (Model II) to provide estimates of the expected number of graduates from Minnesota public high schools in year "t." In contrasting estimates obtained by the two above described methods with estimates independently derived by the Minnesota State Department of Education, it was noticeable that the Department of Education's estimates were substantially lower than the estimates made by the live birth and transition methods for the ten-year period 1960-1970. Since actual data were not available at the time of publication of the report, it was not possible to evaluate the accuracy of any of the models.

Webster (1970) reported the results of a study which determined validity of enrollment projections obtained under the cohort-survival ratio method and the regression approach devised by Brown and Savage. Utilizing data available for a five-year period for 25 Michigan school districts stratified by the factor of student enrollment growth rate, 13 survival ratios for each of the transitional grade-to-grade progressions were calculated and then utilized to predict enrollment for elementary and secondary grades for five additional years.

The multiple regression approach simply developed separate prediction equations for each grade level. The enrollments predicted for elementary grades (1-8) and secondary grades (9-12) were then summed to give separate elementary and secondary total enrollment estimates for each method. A ratio of the difference between predicted and observed enrollment divided by observed enrollment was calculated

as an index of goodness of fit. By noting that regression analysis provided a better estimate of enrollment in 18 of 25 cases, an outcome statistically significant at .05 level, it was concluded that regression analysis was the superior method. It should be noted that Webster's regression procedure paralleled the cohort-survival approach in that regression equations were calculated separately within grade. It would seem that a single multiple regression equation could provide the same degree of accuracy by utilizing birth information to predict total elementary and/or total secondary enrollments.

Multiple Regression Method

More complex regression models with two or more parameters have also been used in order to obtain future enrollment estimates. A description of the rationale and development of models using these regression equations is presented in the following section.

Regression analysis was used by Haggstrom (1969) to develop, specifically for higher education, estimates of demand, future enrollment, and costs as part of a study of the effects of national policy decisions on enrollment in higher education. Haggstrom first established the fact that the logistic growth equation provided an adequate fit to high school graduate and college undergraduate and graduate enrollments for both males and females over the period of time 1947 to 1968. A straight line was fitted to enrollments for the years 1955 to 1968 to obtain estimates of the two parameters required in the logistic growth equation.

At this point the product of the projected high school graduation rates, as derived from the growth curve equation and the age-group projections obtained from the Bureau of the Census, were used to generate projections of high school graduates. Haggstrom noted that undergraduate enrollment rates, except for men during periods of high draft calls and return of veterans to civilian life, showed a steady increase over time. Thus, his function for the projection of female undergraduate enrollment for year t was a nonlinear regression equation that was the product of the estimated enrollment rate as represented by logistic growth equation and the number of high school graduates for year t to t - 3. For the males, the above equation was modified to include coefficients associated with the numbers of veterans under the GI Bill for the last major wars and the number of draftees from the Korean and Vietnam hostilities in a given year t. These terms were introduced to account for the disturbances observed in the original projections of undergraduate enrollment for men.

A second phase of the Brown and Savage (1960) study was concerned with predicting attendance at the University of Minnesota. Five specific models were developed within a multiple regression framework which utilized as input the projections of high school graduates developed for the first phase of the Minnesota educational attendance study.

The first model (Method I) used a least squares approach to fit a time series function to available data for university enrollments for the years 1921-1960. Method II predicted university attendance from a knowledge of the numbers of high school graduates in each year. Estimates of the model parameters were estimated from knowledge of the number of Minnesota high school graduates per year. Inspection of the average squared difference terms obtained from the use of first two models suggested that both models predicted about equally well. However, the sequential runs of negative and positive residuals suggested that at least one important independent variable was not included in these two models. Since it was noted that the greatest discrepancies in prediction occurred during the years of World War II and the Korean War, Method II was extended to take into account national military manpower requirements changes. Evaluation of the new model indicated a smaller average-squared difference (i.e., S^2) when compared to Method II, but poorer predictions were obtained for the years 1921-1940 and 1950-1960. Thus, it was decided that in spite of the decrease in S^2 attained by the use of Method III, Method II was accepted as more efficient since it required less information to obtain roughly the same level of accuracy in enrollment estimates.

The next attempt to refine the prediction process was based upon the recognition that many students do not enter college directly upon graduation from high school; they may work on a job for a year or go into the service. Thus, it was decided to include lagged variables in the model equation. Method IV used a regression equation to predict university attendance as a function of the numbers of high school graduates of the preceding two years and the net changes in military personnel for the preceding four years. It should be noted that Haggstrom arrived at the same general conclusion regarding inclusion of lagged variables and estimates of effects due to war conditions in his model for predicting university attendance in the United States. Method IV showed a good fit over the period from 1921 through 1959 with an S^2 of roughly one-half of that achieved with Method III. Success of Method IV in predicting attendance encouraged the researchers to utilize this same basic approach to generate separate estimates of freshman and other class levels. By aggregating the numbers of freshmen and upperclassmen, it was possible to obtain total institutional estimates. While this procedure would appear to be most likely to provide the best estimate (due to separate equations for estimating freshmen and other students) an observed larger S^2 indicates that Method V was less accurate than Method IV in reproducing the past enrollment data.

The effectiveness of the use of least squares procedures to predict individual course enrollments was demonstrated in the prediction of anticipated course enrollment for educational psychology course at Sacramento State, California, College (Sawhiris, 1970). The first model (Model I) included linear and quadratic time and semester variation (i.e., fall or spring semester) effects to predict attendance in the course. Model II included an effect for overall trend, semester,

and previous semester enrollment as a lagged variable. A third model utilized a polynomial form in an attempt to capture the trend of enrollments over the eight data points available for use in model building. Using the average of the residual squared (i.e., S^2) as a criterion of effectiveness of each of the models to predict attendance, Models II and III were noted to be equally effective while Model I was somewhat less effective. It should be noted that Models I and III used information on sequence of observation while Model II also included information on prior enrollment, a procedure demonstrated by the Minnesota and Haggstrom studies to provide the most efficient prediction in the case of total university enrollment.

The regression procedures featured up to this point are restricted in their output to estimates for one particular phenomenon, be it individual course, total freshman, or total institutional enrollment. This means that if estimates of a different set of phenomena were desired, a different set of equations obtained independently would be required for prediction purposes. The following two sections discuss procedures for generating enrollments estimates that can provide predictions for more than one type of educational phenomena with the same set of equations.

Structural Flow Models

Structural flow models have been used to model student flow through various levels of the educational system. For the purposes of this paper, structural flow models will be defined as models which quantify certain structural relationships among the various factors in the system. Structural models have been widely used to estimate production of doctoral degrees and to study the flow through the system of graduate education in the United States.

Bolt, Koltan, and Levine (1965) developed a structural model which included a feedback element in the procedure to dynamically depict the flow of doctoral degree holders in the fields of science and engineering into various professional activities in and out of higher education. This model was based upon a system of differential equations so that all of the flows, including the so-called feedback flows, were described by two linear difference equations which were simultaneously solved for various values of the equation parameters.

In an extension of the Bolt, et al. model, Reisman (1965) developed a model to encompass the four higher educational production sectors--undergraduate, master, doctoral, and post-doctorate level degrees. Reisman utilized the social-systems simulation methodology developed by J. W. Forrester to solve differential equations interrelating elements in the various equations. This modeling procedure produced estimates of the numbers of graduates expected within the four defined sectors of higher education.

In 1969, Reisman and Taft extended the structure of the 1965 model to include the flow of foreign nationals who study and work within the United States' academic and non-academic systems and to provide explicitly for psychological, sociological, and economic factors which influence the movement of personnel between the levels of the academic and non-academic sectors. This particular model utilized a system of over 200 non-linear difference equations to simulate the production of university degree holders and their feedback into higher education. Difference equations were utilized in this model because of limitations of the available social simulation computer compiler even though it was recognized that the assumption of continuity of student flow was more realistic.

Hammond (1968) proposed a model to depict flow among undergraduate, graduate, post-graduate, academic faculty, and professional non-academic statuses. Four non-linear equations were developed which described the number of individuals within the five groups as a function of an independently determined growth rate and the number of full-time faculty with a Ph.D. or equivalent degree. Hammond used the assumption, as did the previously presented structural models, that the parameters of the model remain constant or change very slowly over time. Utilizing information available on numbers of individuals in the five educational professional categories for all science and engineering areas in the year 1961, Hammond developed estimates for the parameters of the model. These coefficients were used to generate projections for several years into the future. Hammond suggested that the validity of this model rested upon the goodness of fit achieved by his estimates for 1970 and those developed in another study. It should be noted that only Bolt, et al. reported comparisons of model results with actual observed enrollments.

The Organization for Economic Cooperation and Development (OECD) has sponsored the development of models for education utilizing the structural steady state flow concept. While the models discussed above were conceived using modeling student flow so as to obtain estimates of enrollments in various educational categories through the higher education system typical of the kind found in the United States, the interest of the OECD planners has been directed toward modeling of the primary and secondary educational system and the investigation of how this system relates to the economic and/or social subsystems of a particular country.

In 1966, Correa presented a systems model of the elementary and secondary educational structure based upon his work for the OECD. The model utilized information on the number of periods of education offered and on the number of periods of education received in the particular education subsystem being studied in order to generate enrollment estimates at a particular level of disaggregation. While the above model was logically developed, there was no information offered to indicate the validity of the proposed model to project student flow.

Descriptions of a linear "serial flow" and a more general non-linear flow model were presented by Durstine (1969) at the Symposium on Operations Analysis in Education. These models were developed to treat situations that require a model of the flow of students from grade to grade or from one level of an educational system to another in terms of numbers of students residing in different educational states. It was noted that the validity of such flow models is dependent upon the exact definition of categorization or educational segments, the identification of membership in each of the models, and the measurement of flow between the models. Here, also, no information was presented to establish the validity of the proposed model.

Markov-Type Models

The general use of Markov-type flow models based upon transition proportions to describe changes in population distributions over time has been discussed by a number of researchers in the United States and other countries. Before discussing the various types of models utilizing the Markov process approach for predicting student flow, a general introduction to the subject of Markov models will be presented.

In the classical Markov process situation, subjects within the population of interest are distributed into a set of mutually exclusive "states." These states will include the various levels of the educational system under study as well as conditions outside of this system. Some examples of the states would be elementary school, junior high school, senior high school, community college, senior college, graduate school, out-of-school. Parameters of the model are estimated by the obtained transition proportion of the movement of individuals in state i at time t who will be in the same or different state at time t + 1. Time units may be defined as quarter, semester, or single years.

Assuming that the distribution of the population among the states is known for an initial time period t, then the transition matrix can be used to predict the relative frequency within states in a population for a succeeding time, t + 1. This type of model assumes that from tim t to time t + 1, a single individual either remains in his original state or moves to one, and to only one, other state. This movement can be successfully described in terms of the following matrix formulation.

If F_t is defined as an n x 1 column vector of the original distribution of individuals into n states at time t, P is an n x n matrix of transition proportions, and $F_t + 1$ is the n x 1 column vector of n states into which the individuals are categorized at time t + 1, then $F_t + 1 = F_t^T P$. That is, the matrix multiplication of the transpose of input vector F_t by matrix P will give output vector F_t+1. Thus, as the new input entries and transition probabilities are made available for a sequence of time periods, each of equal length, the population

in each of the categories for each of the succeeding time periods may be estimated on the basis of the above matrix multiplications.

It appears that the first application of Markov Chain Theory to model flow through an education system was made by Brown and Savage (1960). As part of their project to test different methods for estimating future enrollment, transition matrices were empirically calculated which described observed student flow into, between, and out of the various curricula within the University of Minnesota for a two-year period, 1957-1959. For the purposes of the study, a student at the University of Minnesota was operationally defined as one who registered, paid fees, and did not withdraw by the end of the second week of the fall quarter of the given year, i.e., either 1957 or 1958. (No distinction was made between full- and part-time students.)

Separate transition matrices were calculated for each sex to describe student flow between colleges and academic classes. The academic classes provided for student classification according to whether the student had freshman through graduate or adult-special status. The transition matrices were utilized jointly with the projections of the numbers of entering students to generate estimates of the numbers of students likely to be in attendance in the various colleges for the fall quarters of 1960 and 1961. The expected enrollments were developed from the regression procedures discussed earlier in this report.

As noted by the authors, the predictions were based upon a single transition matrix, and, thus, no estimates could be made of the stability of the transition probabilities. Therefore, the authors felt the results should be viewed with caution and considered the product of a preliminary investigation of the use of transition matrices to predict college enrollment. However, a visual inspection of the data indicated fair agreement between estimates and actual enrollments for the 1960 fall quarter.

Gani (1963) developed a theoretical model to predict total enrollment and numbers of degrees attained in Australian universities. He used transition matrices of individuals moving from one university level to another to develop estimates of aggregate university enrollments. The model of the university system assumed four undergraduate and three post-graduate years or levels of study. It was further assumed that at the end of each year only three alternatives were available to any student--he could move into the next higher year by passing, he could repeat if he failed, or he could leave the university. Data available for several yars indicated that each transition had a fixed probability. Thus, with the input of the total number of qualified students reaching the age for university entrance, the demand could be predicted for up to 18 years ahead from the known numbers in each age group or cohort. Estimates of the number of bachelor's degrees awarded were found to fit fairly closely the actual number of degrees awarded during a five-year period.

Gani adapted his model to the American university situation in 1965 while in residence at Michigan State University. His "American" university model differed from the Australian model in that it defined progress in terms of credits as opposed to years passed. He also made provisions for transfer between the various schools of the university and for differing student transition rates for the fall, winter, and spring school terms. Unfortunately, Gani did not have a chance to test the validity of his American university model with real data.

A large-scale model utilizing transition proportions to simulate movement within the educational system of Britain and Wales has been developed jointly by the British Department of Education and Science and the Unit for Economic and Statistical Studies on Higher Education (Armitage and Smith, 1967). This large-scale computer model described several different educational states by levels of age. The states included primary school, first-year undergraduate in pure science, primary school teacher, outside world, and deaths. The computer program included a provision for systematic updating of the transition proportions, so that projection could be based upon the most recent available data.

Thonstad (1967) used the theory of absorbing Markov chains to develop a mathematical model of the Norwegian educational system. As noted by the author, the model is based upon the assumptions that a given percentage of pupils enrolled in a certain school will pass their exams successfully and that a certain fraction of these will go to another school while the remainder will take a job. While variables such as school capacities, admission policies, intellectual ability of students, and availability of scholarships do affect the transition ratios, these variables remain consistent enough to allow the model to crudely approximate the educational patterns in Norway. The model provided for 60 different non-absorbing state school activities and 17 unique absorbing states or levels of completed education; death was treated as a separate absorbing state. With this model, one iteration would provide a single year's forecast of school attendance in all parts of the school system, as well as an estimate of the final number of students graduating from the different forms of Norwegian schools.

Another example of the use of Markov chain models is demonstrated by the work of Stone (1965), who attempted to incorporate education as a subsystem into an economic model of Great Britain. Stone's educational subsystem model utilized a discrete time Markov process for graded systems to account for all forms of education (i.e., training and retraining). In Stone's view, the educational system is defined as a system of connected processes where a hierarchy of dependence is formed which accounts for the promotion, retention, and graduation of students. His model contained three sets of parameters: transition rates for flows from one educational process to another, age-specific birth rates, and age-specific death rates. This model is unique in that it utilizes an epidemic-type process. This procedure provided

for yearly changes in transition proportions of students going from one educational level to a higher one as a function of the proportion of individuals that made the transition during the previous year and the proportion that did not make this transition but were academically able to do so. Stone felt the need to extend the model to include intermediate activity levels so that students at the various stages of their academic career could be identified and the requirements for such economic inputs as teachers, buildings, equipment, and supplies could be calculated.

Personnel at the Divisions of Operations Analysis, National Center for Educational Statistics, U. S. Office of Education, have been interested in the use of Markov-type approaches to model student flow. Zabrowski (1968) developed a computerized Markov-type demographic flow model (named DYNAMOD II) to calculate the numbers of individuals in 140 district population groups over selected periods of time.

Separate transition matrices were estimated for each of four sex-race groups across 30 age-educational states with one state to represent deaths. Population data inputs were obtained from the U. S. Bureau of the Census 1/1000 sample data taken from the 1960 Census of Population. Estimates of numbers of births were also obtained from the Bureau of the Census. DYNAMOD II is actually a modified Markov process in that births are included as a separate and variable input at the end of each time interval under consideration. The model euqation states that the number of persons in a particular category in the year t + 1 is equal to the number of persons in that category in year t who remained, plus the number of persons who switch to that category, plus births in the appropriate instances. Zabrowski checked the model against ten-year projections developed in the Office of Education and The Bureau of the Census and concluded that DYNAMOD II gave a reasonable fit to these independent external estimates.

Three separate similation experiments were conducted to demonstrate the feasibility of using a Markov chain flow model like DYNAMOD II to test effects of alternative administrative policy decisions. The experiments were designed to determine the effects of a policy: (1) to increase retention rates of students in the elementary, secondary, and college levels of the national educational system; (2) on student/teacher ratio if retention rates were increased or dropped by a certain percent from what they are at present; and (3) to determine what would be an optimal sequence for implementing educational policies such as increasing retention rates.

Wong (1969a) utilized the Markov chain theory to describe student flow through Columbia University. The purposes of the model were to predict changes in enrollment and movement between divisions that would result from changes in academic policy. Utilizing banked information available for 40,000 students over a period of five years, he calculated transition probabilities for departure rates independent of length of attendance. Information was also obtained to identify significant mobility patterns and lengths of study in a curriculum.

To provide for stable estimates of the parameters of the model, Wong combined the curriculum into six different educational centers or states: undergraduate, which had two centers; Graduate, which had three centers; and Combined or Professional, which had one center. The resultant model was used to simulate movement of students between six centers under three independent constraints of capacity--arrival and departure rates and length of attendance in a center. As a result of experiments utilizing his model to simulate student flow, he found that student movement was a function of university admissions policies and operating procedures.

Two other reports of the use of Markov-type models will be now presented to conclude this section. While not specifically concerned with prediction of enrollment in educational institutions, these two projects were concerned with the descripton of flow of individuals between stages of occupational development which can be considered as analogous to the movement of students between curricula.

One of the first applications of a Markov chain model to model flow of personnel was reported by John Merck in 1959. He reported on the use of a Markov chain model to estimate short-term and long-term effects of policy decisions on the Airman Personnel System. Since that time, Merck and his colleagues at the Personnel Research Laboratory at Lackland Air Force Base, Texas, have reported the results of several studies utilizing the Markov chain theory to investigate questions relating to the maintenance of an adequate number of Air Force personnel to the level necessary for carrying out the role of the Air Force. These studies were concerned with such problems as retention of first enlistment airmen (Merck, 1962), prediction of retirement rates (Harding and Merck, 1964), and projecting movements of personnel through a system (Merck, 1965).

The 1965 personnel model was a sophisticated, computer-processed mathematical model which simulated movements of personnel through the system. In this model, movement was based upon empirically derived transition probabilities. Significant variables such as career fields and service grade were used to define the states of the model. Using future enlistment estimates as inputs, the model was iterated to produce the estimated distribution of personnel at the end of the next time interval.

Lohnes and Gribbons (1970) demonstrated the usefulness of Markov chain models to describe the development of career development over time. Utilizing a career variable with four nominal measurement levels, they investigated the stability of career aspirations over four separate time periods. Of particular importance to this review was the use of procedures developed by Anderson and Goodman (1957) to fit a stationary transition matrix to the last of the empirically derived transition matrices of career development. It was noted that the null hypotheses presented were upheld for the stationary hypothesis; however, the null hypothesis that the transition matrix had a one-step memory was rejected,

thus suggesting that the data did not satisfy the assumptions required by a Markov chain model.

This concludes the review of various types of models utilized to project student flow. The next section will discuss a proposed rationale for selecting a procedure to model the student flow within a state community college system and within specific community colleges.

CONSIDERATIONS IN SELECTING
A MATHEMATICAL MODEL
FOR ESTIMATING ENROLLMENT

A crucial concern in the development of any mathematical model is the availability of well ordered data. The accuracy of any derived model will be constrained by the availability of accurate data. Also, the cost of developing estimates of model parameters will be directly proportional to the accessibility and arrangement of available data. Several researchers have noted this to be a particularly important concern in the development of a Markov-type model. (See, for example, Brown and Savage, 1960, p. 42; Merck, 1965, p. 11-13; Wong, 1969a, p. 11; and Wurtele, 1967, p. 30.)

In general, educational statistical data tend to be aggregated; schools report total enrollments by class, numbers of teachers by subject, etc. An aggregated data system, even when classifications are not highly aggregated, provides information on distributions of students among various educational activities. However, these educational stocks can only provide approximations of the rates of movement between states and, thus, of the flow of students between educational activities. These types of data would appear to provide adequate information for the development of a structural flow model and the extrapolation types of models, but the Markov-type models require actual information on the movement of individual students through their educational career. Thus, aggregated data cannot provide the required estimates of transition probabilities required for the development of a Markov-type model.

A second consideration is whether or not the developed mathematical model is to be based upon a deterministic or probabilistic type process. That is, the model builder can set up a set of equations that either does or does not include probabilistic components.

If the effect of any change in the system can be predicated with certainty, it is said to be deterministic. If not, then a probabilistic component may be included to account for the discrepancies between the predictions made by the model and actual behavioral outcomes. When the model is concerned with a sequence of events where the outcome on each particular event depends upon some chance element, than the sequence is called a stochastic process. Thus, it can be seen that the regression model approach would be considered probabilistic, the structural flow model deterministic, and the Markov chain approach stochastic.

In most attempts at model building probabilistic components are introduced as a matter of the model building strategy selected and not as a function of whether or not the relevant behavioral process under study is really stochastic or derministic.

It should be noted that the use of a deterministic model to make exact predictions of behaviors such as student movement between curricula in post-secondary institutions would have to be extremely complex; an example would be the flow model developed by Riesman, et al, which required over 200 equations to describe flow between post-secondary educational levels.

The present paper developed out of a project concerned with the projection of enrollment in and student flow between the various curricula offered in a community college. It would seem that the career plans of community college students are not completely fixed, thus leading to a situation where student flow within an institution cannot be predicted with complete accuracy. While it may be that information relating to elements necessary for accurate prediction of a student's future educational or work plans cannot be obtained, it would seem that a stochastic model based upon a Markov chain-type process would be the appropriate method to utilize in generating estimates of student enrollment in various curricula over time. Also, attempts to predict total institutional enrollment would seem to be best accomplished by a probabilistic model. Thus, regression models would appear to be appropriate for a single enrollment estimate as would be required for a single institution or for the state for a particular year.

In commenting upon an earlier proposal to model student flow through a community college, Wong (1969b) noted that structural flow models tend to oversimplify the movements of students within the various levels of the educational system. He further pointed out that the use of continuous distributions of movements over time but independent with respect to time over a wide range of values are necessary for the development of a structural flow model. Since students "travel" indiscrete steps through the present educational system, he concluded that the use of a Markov-type process would provide a more accurate model of student movement within a community college system than would a structural flow model.

The theory of Markov chains assumes that there is a constant transition probabilities matrix for the population from which each of the observed empirical transition matrices was sampled under the constraint that they may be subject to random sampling errors. A second assumption is that the probabilities of the various outcomes for a subject at any transition is based upon his status at the prior time period and not on his status more than one time period removed. These two assumptions are referred to as the stationary hypothesis and the one statistical step dependency hypothesis and can be tested for significance by procedures developed by Anderson and Goodman (1957). Lohnes and Gribbons (1970) seriously questioned this one-state dependency assumption in the case of career aspiration. However, they were able to find support for the stationary hypothesis (the first assumption required for Markov chain models.) It has been noted by Billingsley (1961) that no natural (e.g., social) process exactly satisfies the Markov chain condition, but many will come close enough to make a Markov chain model useful. Creager (1970) notes that when the requirements for a Markov chain model are

relaxed, greater flexibility and realism in reflecting the educational process accrue. While this means the extensive mathematical development associable with the classical Markov model is no longer applicable, the multiplicative relationship between input and output distributions of students in the educational states still holds. That is, $F_2 = F_1^T P$ can still be calculated since these are simple transition matrix calculations not dependent upon Markov chain assumptions. The case of $P_{13} = P_{12}^T P_{13}$ is an example of several probability transition matrices being reduced from multi-panel data to a single one-stage overall transition matrix.

In this case, a stationary probability matrix as required for classical Markov chain theory between stages does not need to be assumed. While many stochastic models do search for a stationary transition matrix as a "steady-state," it does not appear that this is a relevant question to ask of the movement of students through a post-secondary education system. Based upon the above discussion, it seems apparent that a model based upon a transition matrix will provide the most efficient approach to developing enrollment estimates in various levels of an educational system. While the transition matrix as generated from empirical data may not meet the assumptions of a classical Markov-type process, the obtained transition proportion matrix can still be used in the matrix multiplication approach suggested by Creager to obtain estimates of the distribution of students in educational states of a defined system.

The above discussion has been oriented toward presenting a variety of approaches that can be used by educators in developing enrollment projections. However, it is also appropriate to acknowledge that some individuals have reservations about the appropriateness of the application of modeling procedures to education.

Alper (1968), in particular, suggests that the application of the systems analysis procedures to educational planning models has been unsuccessful, and he proceeds to enumerate seven difficulties in such attempts. He feels that much of the difficulty arises from the use of input-output analysis in the development of educational models. Of the seven concerns noted, two appear to be relevant to the above discussion. The first (Alper, #2, p. 95) is specifically concerned with the use of deterministic models without probabilistic components to model educational activities, and the second (Alper, #5, p. 95) is concerned with the assumption that model parameters will remain constant over time. A discussion of the drawbacks of the use of deterministic models has already been presented in connection with why this study does not propose to develop a structural flow model to describe the movement of students between curricula in a community college. As for the second concern, the best estimate one can make for the future is that it will follow the trends already evident to the researcher. However, it is also acknowledged that if substantial change in educational administrative policy or a change in perception of the desirability of schooling were to occur, then model parameters would have to be revised to take into account the observed changes. This possibility suggests the need for a constant updating of

information utilized to generate enrollment projections. It can be stated that the projections developed from a Markov-type model will be valid to the extent that the model parameters have the same values as they did at the time the original model was developed.

CONCLUDING REMARKS

The purpose of the present report was to discuss the use of methods for generating enrollment projections. An attempt was made to develop a scheme for categorizing the various types of models reported in the educational planning. A cursory inspection of the dates of the published reports referred to in this paper indicates that enrollment projections utilizing methods other than the cohort-survival ratio procedure are of recent origins. In fact, one could conclude that the development of more complex mathematical models has paralleled the increasing availability to researchers of electronic computers.

While this document attempted to do an exhaustive survey of existing methods, it is quite likely that other procedures which have not found their way into print have been developed and successfully used in projection of school and/or class enrollment. However, it is felt that the present report does provide an extensive listing of the types of procedures which could be used to provide estimates of future educational demands at various levels of the educational system.

To recapitulate, the types of models presently available for projection purposes can be considered essentially demographic educational systems models. The models presently used in educational enrollment forecasting were also categorized by type of procedure used, i.e., straight-line or extrapolation, structural, or Markov-type. It was concluded that models of the first type using a regression form were likely to provide the most efficient means for singular estimates of enrollments such as would be required for a single institution. In contrast, where enrollments for a number of differentiated parts of the same system are needed, it was concluded that a Markov-type model would likely provide the most efficient estimates of differential educational demand.

REFERENCES

Adair, C. D. "Predicting New Year's Enrollment." School and Community, 55, 1969, 37.

Alper, P. "A Critical Analysis of the Application of Systems Analysis to Educational Planning Models." IEEE Transactions in Education, E-11, 1968, 94-98.

Anderson, T. W. and L. A. Goodman. "Statistical Inference about Markov Chains." Annals of Mathematical Statistics, 28, 1957, 89-110.

Armitage, P. and C. Smith. "The Development of Computable Models of the British Educational System and their Possible Uses." Mathematical Models in Educational Planning. Paris, France: Directorate for Scientific Affairs, Organization of Economic Co-operational Development, 1967.

Billingsley, P. "Statistical Methods in Markov Chains." Annals of Mathematical Statistics, 32, 1961, 12-40.

Bolt, R. H., W. L. Koltan, and O. H. Levine. "Doctoral Feedback into Higher Education," Science, 148, 1965, 918-928.

Brown, B. W. and J. R. Savage. Methodological Studies in Educational Attendance Prediction. Minneapolis, Minnesota: Department of Statistics, University of Minnesota, 1960.

Correa, H. "Basis for the Quantitative Analysis of the Educational System." Journal of Experimental Education, 35, 1966, 11-18.

Creager, J. A. Use of Empirical Transition Matrices in Educational Research. Paper read at American Educational Research Asssociation Annual Meetings, Minneapolis, Minnesota, March, 1970.

Cross, R. H. and C. H. Sederberg. "Computer-Assisted Enrollment Projections Procedures." Journal of Educational Data Processing, 2, 1969, 160-165.

Durstine, R. M. "In Quest of Useful Models for Educational Systems." Socio-Economic Planning Sciences, 2, 1969, 417-437.

Feller, W. An Introduction to Probability Theory and its Applications, Vol. 1 (2nd ed.). New York: John Wiley, 1957.

Gani, J. "Formulae for Projecting Enrollments and Degrees Awarded in Universities." Journal of the Royal Statistical Society, Series A, 126, Part 3, 1963, 400-409.

Gani, J. A Model for Student Enrollment in American Universities: I. Structure of the Model. Unpublished manuscript, 1965.

Haggstrom, G. W. On Analyzing and Predicting Enrollments and Costs in Higher Education. Paper read at American Statistical Association, New York, August, 1969.

Hammond, A. A Flow Model for Higher Education. Santa Monica, California: Rand Corporation, August, 1968.

Harding, F. D. and J. W. Merck. Markov Chain Theory Applied to the Prediction of Retirement Rates. Technical Documentary Report PRL-TDR-64-14. 6570th Personnel Research Laboratory, Aerospace Medical Division, Air Force Systems Command, Lackland Air Force Base, Texas, June, 1964.

Koulourianos, D. T. Educational Planning for Economic Growth, Technical Report 23, Center for Research in Management Science, University of California, Berkeley, California, February, 1967. In P. Alper, "A Critical Analysis of the Application of Systems Analysis to Educational Planning Models," IEEE Transactions on Education, E-11, 1968, 94-98.

Lohnes, P. R. and W. D. Gribbons. "The Markov Chain as Null Hypothesis in a Development Survey." Journal of Educational Measurement, 7, 1970, 25-32.

Merck, J. W. Retention of First Enlistment Airmen: Analysis of Results of a Mathematical Simulation. Technical Documentary Report PRL-TDR-62-17. 6570th Personnel Research Laboratory, Aerospace Medical Division, Air Force Systems Command, Lackland Air Force Base, Texas, March, 1965.

Merck, J. W. and F. B. Ford. Feasibility of a Method for Estimating Short-Term and Long-Term Effects of Policy Decisions on the Airman Personnel System. Technical Report WADC-TR-59-38. Personnel Laboratory, Wright Air Development Center, Air Research and Development Command, United States Air Force, Lackland Air Force Base, Texas, 1959.

Reisman, A. "Higher Education: A Population Flow Feedback Model." Science, 153, 1966, 89-91.

Reisman, A. and M. I. Taft. "The Generation of Doctorates and Their Feedback into Higher Education." Socio-Economic Planning Sciences, 2, 1969, 473-486.

Sawhiri, M. Y. "The Projection of College Enrollment." Multivariate Behavioral Research, 5, 1970, 83-116.

Stone, R. "A Model of the Educational System." Minerva, 3, 1965, 172-186.

Thonstad, T. "A Mathematical Model of the Norwegian Educational System." In Mathematical Models in Educational Planning. Paris: Organization for Economic Co-Operation and Development, 1967, 125-158.

Webster, J. "The Cohort-Survival Patio Method in the Projection of School Attendance." Journal of Experimental Education, 39, 1970, 89-96.

Wong, Y. "Computer Simulation of Student Mobility Patterns." Journal of the Association for Educational Data Systems, 2, 1969, 3-19.

Wong, Y. Personal communication, 1969.

Wurtele, Z. S. Mathematical Models for Educational Planning. Professional Paper. SP-3015. System Development Corporation, Santa Monica, California, 1967.

Zabrowski, K. "The Dynamod Model of Student and Teacher Population Growth." Socio-Economic Planning Sciences, 2, 1969, 455-464.